



Disentangled Representation Learning for Leaf Diseases Recognition

Xing Wang, Congcong Zhu, and Suping Wu^(✉)

School of Information Engineering, Ningxia University, Yinchuan 750021, China
wx_nxu@163.com, ccz_nxu@163.com, wspg123@163.com, pswuu@nxu.edu.cn

Abstract. Plant disease detection plays an important role in agricultural production and ecological protection. However, it is always a challenge to detect the severity of plant diseases in multi-species and multi-disease conditions. Unlike most existing classification methods which are difficult to solve multi-properties detection, we propose a disentangled representation interactive network (DRIN), which disentangles the global features of each plant leaf and learns the discriminative representation of multiple sub-properties, including plant species, disease types and disease severity. To achieve it, the disentangled representation network transform the joint probability into the conditional probability through the information interaction between the sub-properties. Moreover, data filtering was introduced to reduce the error messages in property interactions. Experimental results demonstrate the effectiveness of our DRIN on the plant disease detection dataset.

Keywords: Representation learning · Plant disease detection · Information interaction · Data filtering

1 Introduction

With the continuous development of computer technology, the research on intelligent identification of plant diseases has made good progress. The main challenge existing is the fact that it can only detect specific plant diseases or multiple plant diseases, but cannot distinguish the severity of the disease. It is important to distinguish the severity of the disease, for it represents different strategies in dealing with the disease (e.g. the medicine quantity).

There are many traditional methods for detecting plant diseases. Tan et al. [1] established a multi-layer BP neural network model by calculating the leaf chromaticity value to realize the identification of soybean leaf diseases. Tian et al. [2] used the support vector machine (SVM) recognition method to extract the color and texture features of grape diseased leaves, and achieved a better recognition effect than the neural network. Wang et al. [3] extracted the color, shape,

This work was supported in part by the Ningxia Key Research and Development Plan Major Projects under Grant 2016BZ0901 and in part by the Research and Innovation Foundation of First-class Universities in Western China under Grant ZKZD2017005.

texture and other features of leaf lesions, combined with environmental information, and used discriminant analysis method to identify the types of cucumber lesions. Zhang et al. [4] also extracted the color, shape and texture features of the diseased spots after the spots were segmented, and then identified five kinds of corn leaves by the k-nearest neighbor (KNN) classification algorithm. In the literatures above, specific plant image features were extracted and combined with traditional classification methods to identify diseases. Although the methods above have achieved good recognition effects, they are not able to completely represent plant disease information due to specific features. For instance, some certain diseased leaves may appear in other features (e.g. powder) rather than disease spots, which makes the segmentation more difficult and has a negative impact on the recognition effect. Moreover, the number of test samples selected by the methods above is limited, i.e. the selected leaves are only from one plant, and these methods are only limited to the identification of the same plant leaf diseases.

In recent years, convolutional neural network [5] has been widely used in the field of image recognition (such as handwritten font recognition [6], face recognition [7,8] and object detection [9,10]) without relying on specific features. In general recognition, convolutional neural network models such as AlexNet [11], GoogLeNet [12] and ResNet [13] have achieved good results. An increasing number of scholars applied these models to image recognition in a narrow sense. [14,15] used convolutional neural network to conduct relevant studies on plant leaf classification. Sladojevic et al. [16], Brahimi et al. [17] and Amara et al. [18] applied the convolutional neural network to the identification of plant leaf diseases, and improved the model CaffeNet and AlexNet respectively with fine-tuning methods, achieving good identification results. The literatures above proved that convolutional neural network is feasible to identify plant leaf diseases. In the field of agriculture, it is far from enough to just identify the type of plant diseases, which did not consider collaborative predictions of species, disease types and disease severity. Therefore, modeling the interaction between multi-properties is critical.

To address the above-mentioned challenges, in this paper, we propose a disentangled multi-representation interactive network to identify various plant diseases. Different from the existing detection methods of plant diseases, which can only detect one or more diseases, the DRIN adopts the learning methods of property decomposition, property interaction, property fusion and data filtering, aiming at predicting the plant disease severity and improving classification accuracy. Specifically, the DRIN disentangle the total plant property into three sub-properties: species, disease types and disease severity. In the process of sub-property prediction, the interactive information was added between sub-properties and fuse the information of sub-property species into sub-property disease and disease severity respectively. At the same time, we integrate the information of sub-property diseases into the severity of sub-property diseases. After that, we fuse the information of multiple sub-properties into the final result by means of property fusion. Finally, in order to ensure the effective information

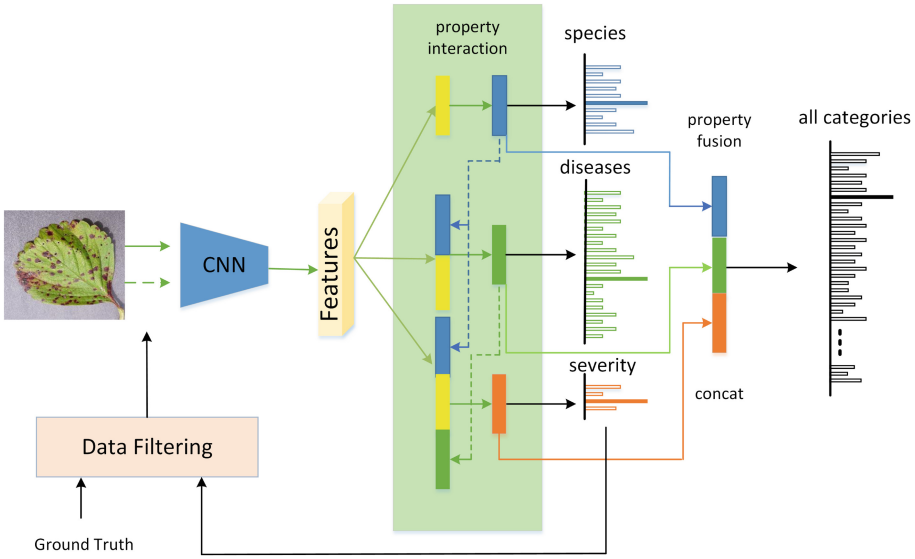


Fig. 1. Schematic diagram of the network structure of our method. We first disentangle the total plant property into three sub-properties: species, disease types and disease severity. In the process of sub-property prediction, we add the interactive information between sub-properties and fuse the information of sub-property species into sub-property disease and disease severity respectively. At the same time, we integrate the information of sub-property diseases into the severity of sub-property diseases. After that, we fuse the information of multiple sub-properties into the final result by means of property fusion. Finally, in order to ensure the effective information transfer in property interaction, we introduce data filtering to reduce error information in the process of property interaction.

transfer in property interaction, data filtering was introduced to reduce error messages in the process of property interaction. Figure 1 shows the structure of our framework.

The core contributions are summarized as follows:

- For the first time, we use feature learning in plant disease prediction. To achieve it, multi-branch network was proposed to disentangle the global information of plant leaves. Therefore, this method avoids the problem of difficult optimization in joint prediction of multiple sub-properties.
- We propose a property interaction to change the joint probability into the conditional probability through the information interaction between sub-properties. For error information, data filtering was introduced to reduce it in the process of property interaction.

2 Our Approach

In this paper, multi-branch network was introduced to disentangle the global information of plant leaves. This method avoids the problem of difficult optimization in joint prediction of multiple sub-properties. In addition, the DRIN put forward the property interaction method to transform the joint probability into conditional probability. Furthermore, a data filtering was introduced to reduce the property interaction error messages.

2.1 Multi-properties Interactive Networks

Detection of disease severity is a complex classification problem in multi-species and multi-disease conditions, and the network needs to be able to simultaneously predict species, disease and disease severity. The traditional network can only detect these three properties jointly, which leads to increase fitting complexity of the model. In order to solve this problem, we transform the joint prediction task into multiple sub-property classification tasks.

Given an input image I , we first put the image into the convolutional layer of pre-training to extract the features of plant diseases. The extracted deep representations are denoted as $F = W_c * I$, where $*$ denotes a set of operations of convolution, pooling and activation, and W_c denotes the overall parameters. Then the extracted global features are sent to the disentangle representational network. In the disentangle representational network, the three branch networks learn the higher-order feature expressions of species, disease types and disease severity respectively. This clearer expression of higher-order features contributes to more accurate detection tasks. Finally, the three sub-properties are fused to obtain the final result. The extracted feature F obtained species feature F_s , disease feature F_d and disease severity feature F_l through species decomposition function f_s , disease types decomposition function f_d and disease severity decomposition function f_l , respectively. Therefore, the formula to formalize the property decomposition is as follows:

$$F_s = f_s(F), F_d = f_d(F), F_l = f_l(F) \quad (1)$$

The disentangled representation sub-network has a natural hierarchical relationship among multiple sub-properties. For example, it's easier to identify a disease with a known species. Similarly, it is easier to deduce the severity in the presence of a known species and disease. There is no doubt that this natural hierarchy can provide prior information in multi-properties tasks. Therefore, we make use of the probabilistic formulation over the variables including the image I , species S , disease D and severity of plant disease L . In accordance with the natural relation of the interaction between multiple sub-properties, the joint probability and conditional probability under multiple properties can be written as:

$$p(S, D, L, I) = p(L|S, D, I)p(D|S, I)p(S|I)p(I) \quad (2)$$

$$p(S, D, L|I) = \frac{p(S, D, L, I)}{p(I)} = \underbrace{p(S|I)}_{\text{Species}} \cdot \underbrace{p(D|S, I)}_{\text{Disease}} \cdot \underbrace{p(L|S, D, I)}_{\text{Severity}} \quad (3)$$

We have carried out the information interaction between the sub-properties in the disentangled representation sub-network which approximately transforms the joint probability into conditional probability.

After the property decomposition module, there is a very important fusion module, which fuses the results of decomposing sub-properties to obtain the final predicted results. There are intersections and differences between species and disease (for example, two widely different species may share a disease, and two similar species may not have the same one). Thus, there is a possibility that the species sub-network and disease sub-network predict two outcomes those are absolutely impossible to coexist (like citrus is free from powdery mildew). Therefore, we introduce the concept of mutual supervision to constrain sub-properties fusion. When two incompatible results appear, the result with the highest confidence in two properties is taken as the prediction guide. In the other sub-properties, a high confidence result that coexists with the previous sub-property is selected as the prediction result. By using the mutual supervision information among the sub-properties, the dependence among the branch networks can be well restrained.

2.2 Data Filtering

One of the most challenging aspects of this task is the intersection of species, disease and severity. There are difficult samples that are similar in certain property. As shown in Fig. 2, it can be seen that there are hard samples in the data set that cannot be distinguished by people. This means that when some samples are highly similar in one property, it will confuse the judgment in other sub-properties. For example, when the sample prediction is wrong, it means that the condition is wrong in the conditional probability of the disentangled representation network, so the network will predict the probability under the wrong condition. The two problems above will affect the gradient descent direction of the network and make it difficult for the network to converge to the optimal value.

To solve this problem, we propose a method of data filtering. As shown in Fig. 3, the network will conduct two iterations. The first iteration uses all the data in batch size to train the network. After that, the data will be filtered once and the network gradient direction under correct conditional probability will be optimized again.

During the data filtering phase, a batch size sample disease severity label was used to test the consistency with the predicted results. When the label is consistent with the predicted result of the network, the sample will be fully used to train the network. When the label is inconsistent with the network prediction result, the confidence of the positive class in the prediction result will be taken as



Fig. 2. The pictures in the first line show the general plant diseases severity. The second line shows the severity of the plant disease. From left to right are Citrus Greening June, Pepper Scab, Grape Leaf Blight Fungus and Peach Bacterial Spot.

the weight value of loss calculation for this sample. When the label of a sample is significantly different from the predicted result of the network, the influence of it as an error condition on the classification in other sub-properties will be greater. Therefore, it will produce false conditional probability to affect the prediction accuracy. In the second iteration, our network reduce the contribution of this sample to network optimization. Through data filtering, data will describe the difference between properties more accurately, and more robust data distribution will guide the conditional probability more accurately between sub-properties.

3 Experiments

To evaluate the effectiveness of the DRIN, we firstly conducted the main experimental results on the 2018 Global AI Challenge Plant Disease Datasets¹. Next, we present details on evaluation data set, protocol and experimental analysis, respectively.

3.1 Evaluation Dataset and Protocol

The dataset has 61 classifications (according to “species-disease-degree”), 10 species, 27 diseases (24 of which are classified into general and severe), 10 health classifications, and a total of 47,393 pictures. Each picture contains a leaf of one plant, and the leaf occupies the main position of the picture. The dataset is randomly divided into four sub-data sets: training (70%), validation (10%), test A (10%) and test B (10%). Among them, the training set has 32,739 pictures,

¹ <https://challenger.ai/competition/pdr2018>.

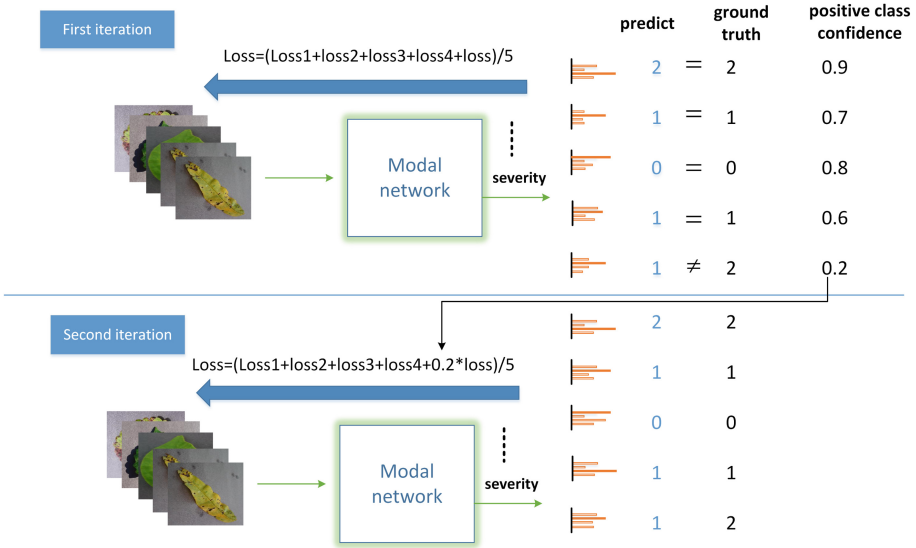


Fig. 3. Schematic diagram of data filtering structure. In the data filtering phase, the batch size is assumed to be five. When the five photos are iterated for the first time, the final loss value is equal to the average value of the sum of each loss. When the predicted result of the first iteration is inconsistent with the label, the loss of the second iteration is multiplied by the confidence of the predicted result of the positive class; When the predicted results of the first iteration are consistent with the label, the loss of the second iteration remains unchanged.

the validation set has 4,982 pictures, the test set A has 4,959 pictures, and the test set B has 4,957 pictures. Since the labels for test set A and test set B are not publicly available in the dataset, We mix the two data sets of training and validation, and finally randomly select 10% of them as the test set and the rest as the training set.

3.2 Implementation Details

Datasets are unevenly distributed across the data distribution. In the training dataset, there are only two images of tomato scab, one general and one serious, which cannot be trained to achieve good result, so we delete these two images. For other data, the training set was perturbed by randomly rotation, horizontal and vertical flip to increase the data to solve the problem of data imbalance. At the same time, we scale the image to 224×224 pixels for training. For optimization, Adam optimizer was used with a learning rate begins at 0.0001 and decays 0.9 after each 20 epochs. The batch size is set as 40.

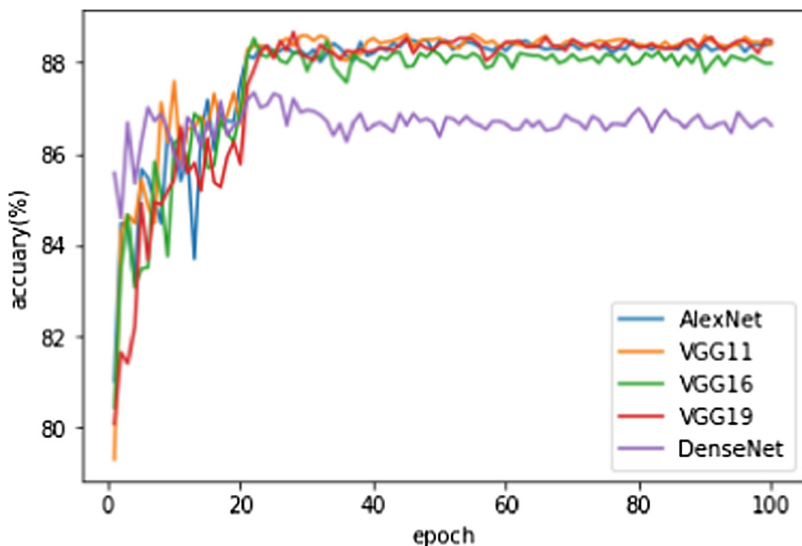


Fig. 4. The accuracy curve of classical network architecture in plant disease data set. From these results, the performance of VGG19 surpasses that of other classic network structures.

3.3 Results and Analysis

Plant disease dataset was trained in the model DRIN. To evaluate the methods, we use the classical network structure Alexnet, VGG11 [20], VGG16, VGG19 and DenseNet [19] to classify all the categories. The accuracy curve shown in Fig. 4 is obtained. From these results, the performance of VGG19 surpasses that of other classic network structures.

In order to prove the effectiveness of the method, we select VGG19 as our basic network and design three groups of experiments. Specifically, we put an image into the VGG19 network, and then enter three different full-connection layers to decompose its property into three sub-properties (species, disease types, and disease severity). There are interactions between sub-properties in the full-connection layers, and finally the three sub-properties was fused to obtain the final result. At the same time, in order to prove the effectiveness of the data filtering, we design two sets of experiments. Specifically, The former is the basis of the first set of experiments. We cancel the interaction between sub-properties of full connection and increase data filtering. The latter is based on the first set of experiments, the way of data filtering was added. The prediction accuracy of experimental results is shown in Table 1.

As can be seen from Table 1 that VGG19 has the best performance in the classical deep learning network structure, which is far superior to other deep learning network structures. The experimental results show that when property decomposition, property interaction and property fusion are used in the method, the performance is improved by 0.08%, which proves the effectiveness

Table 1. Comparisons of accuracy of the proposed DRIN with classical network framework. PD, DF, PI and PF respectively represent property decomposition, data filtering, property interaction and property fusion. We found that our approach achieved better performance than classical network structures.

Methods	Accuracy(%)
AlexNet	88.42
DenseNet	86.62
VGG11	88.39
VGG16	87.92
VGG19	88.47
PD+PI+PF(ours)	88.55
PD+DF+PF(ours)	88.44
PD+PI+DF+PF(ours)	88.86

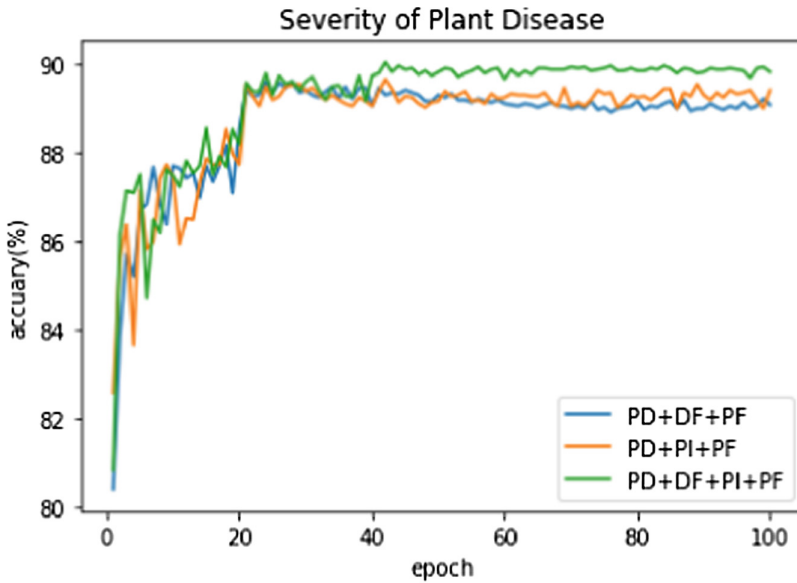


Fig. 5. The curve of the accuracy of each group in predicting the severity of plant diseases was shown. PD, DF, PI and PF respectively represent property decomposition, data filtering, property interaction and property fusion. We found that our methods with PD, DF, PI, and PF performed better than others.

of the property interaction in our experiment. When adding data filtering to the method, the performance is improved to 88.86% and increased by 0.29%. At the same time, we do a counter example experiment in which data filtering is added when there was no property interaction in the experiment, so the accuracy of the experiment was reduced. This experiment proves that the proposed data filtering

method reduces the weight of error message interaction in property interaction, and thus proves the effective performance of data filtering. Since it is the most difficult part to predict the severity of diseases in the sub-property results, we have drawn the curve of the prediction accuracy of plant disease severity, as shown in Fig. 5.

4 Conclusion

We propose a disentangled representational interactive network to solve the problem of predicting plant disease in the case of multiple plants and multiple diseases. Our method consists of property decomposition, property interaction, property fusion and data filtering. Property decomposition is to disentangle the entire property into three sub-properties. Property interaction refers to the information transfer between sub-properties. Property fusion is the fusion of three sub-properties into final result. Data filtering reduces the transmission of error messages between sub-properties. Experimental results demonstrate the effectiveness of our method for plant disease detection. In future work, we will use metric learning to further improve our performance.

References

1. Feng, T., Xiaodan, M.: The method of recognition of damage by disease and insect based on laminae. *J. Agric. Mechan. Res.* **6**(2009), 41–43 (2009)
2. Youwen, T., Tianlai, L., Chenghua, L., Zailin, P., Guokai, S., Bin, W.: Method for recognition of grape disease based on support vector machine. *Trans. Chin. Soc. Agric. Eng.* **23**(6), 175–180 (2007)
3. Wang, X., Zhang, S., Wang, Z., Zhang, Q.: Recognition of cucumber diseases based on leaf image and environmental information. *Trans. Chin. Soc. Agric. Eng.* **30**(14), 148–153 (2014)
4. Zhang, S., Shang, Y., Wang, L., et al.: Plant disease recognition based on plant leaf image. *J. Anim. Plant Sci.* **25**(3 Suppl. 1), 42–45 (2015)
5. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014*. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
6. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al.: Gradient based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
7. Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1891–1898 (2014)
8. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9911, pp. 499–515. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_31
9. Ren, S., He, K., Girshick, R., Sun, J.: RCNN Faster. [n.d.]. Towards Real-Time Object Detection with Region Proposal Networks ([n. d.]

10. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587 (2014)
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
12. Szegedy, C., et al.: Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9 (2015)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
14. Gong, D.X., Cao, C.R.: Plant leaf classification based on CNN. *Comput. Modernization* **4**(2014), 12–15 (2014)
15. Zhang, S., Huai, Y.: Leaf image recognition based on layered convolutions neural network deep learning. *J. Beijing For. Univ.* **38**(9), 108–115 (2016)
16. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D.: Deep neural networks based recognition of plant diseases by leaf image classification. *Comput. Intell. Neurosc.* **2016**, 11 (2016)
17. Brahimi, M., Boukhalfa, K., Moussaoui, A.: Deep learning for tomato diseases: classification and symptoms visualization. *Appl. Artif. Intell.* **31**(4), 299–315 (2017)
18. Amara, J., Bouaziz, B., Algergawy, A., et al.: A deep learning based approach for banana leaf diseases classification. In: *BTW (Workshops)*, pp. 79–88 (2017)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)* (2014)
20. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708 (2017)